A statistical-mechanical study of evolution of robustness in noisy environment

Ayaka Sakata* and Koji Hukushima†
Graduate school of Arts and Sciences, The University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan.

Kunihiko Kaneko‡
Graduate school of Arts and Sciences, The University of Tokyo, Komaba, Meguro-ku, Tokyo 153-8902, Japan.
Complex Systems Biology Project, Exploratory Research for Advanced Technology (ERATO), Japan Science and Technology Agency (JST), Tokyo, Japan

(Dated: September 8, 2009)

In biological systems, expression dynamics that can provide fitted phenotype patterns with respect to a specific function have evolved through mutations. This has been observed in the evolution of proteins for realizing folding dynamics through which a target structure is shaped. We study this evolutionary process by introducing a statistical-mechanical model of interacting spins, where a configuration of spins and their interactions $\boldsymbol{J}$ represent a phenotype and genotype, respectively. The phenotype dynamics are given by a stochastic process with temperature T_S under a Hamiltonian with $\boldsymbol{J}$. The evolution of $\boldsymbol{J}$ is also stochastic with temperature T_J and follows mutations introduced into $\boldsymbol{J}$ and selection based on a fitness defined for a configuration of a given set of target spins. Below a certain temperature T_S^{c2}, the interactions $\boldsymbol{J}$ that achieve the target pattern evolve, whereas another phase transition is observed at $T_S^{c1} < T_S^{c2}$. At low temperatures $T_S < T_S^{c1}$, the Hamiltonian exhibits a spin-glass like phase, where the dynamics toward the target pattern require long time steps, and the fitness often decreases drastically as a result of a single mutation to $\boldsymbol{J}$. In the intermediate-temperature region, the dynamics to shape the target pattern proceed rapidly and are robust to mutations of $\boldsymbol{J}$. The interactions in this region have no frustration around the target pattern and results in funnel-type dynamics. We propose that the ubiquity of funnel-type dynamics, as observed in protein folding, is a consequence of evolution subjected to thermal noise beyond a certain level; this also leads to mutational robustness of the fitness.

PACS numbers: 87.10.-e,87.10.Hk,87.10.Mn,87.10.Rt

I. INTRODUCTION

The function of a biological unit is generally determined by a phenotype, which is the result of a dynamical process that yields a specific pattern or structure. The dynamics themselves are governed by a genetic sequence. In evolution governed by a fixed fitness condition, a phenotype that gives a higher value for the fitness function is selected. The genes that produce such a phenotype are transferred to the next generation, and thus, specific genetic sequences are selected. However, we note, that the dynamical process producing a phenotype is subject to noise. Accordingly, the genotype-phenotype mapping is generally stochastic.

For example, genetic information determines the amino acid sequence for a protein, while the tertiary structure responsible for its functions is generated only by folding dynamics. By the folding process, a structure is formed in the protein, and this structure serves as the basis for the function. The genotype-phenotype mapping is formed by this folding process, and this mapping is stochastic because of the thermal noise in the folding process[1]. Related folding dynamics also occur in t-RNA, where the influence of thermal noise on genotype-phenotype mapping has been intensively investigated[2]. At a more macroscopic level, genetic information specifies a gene regulatory network, which determines the dynamics for the gene expression pattern, thus giving rise to the phenotype. This gene expression dynamics are again stochastic because the number of proteins in a cell is not necessarily very large[3]. In fact, the stochasticity for gene expression of isogenic organisms has been studied extensively[4, 5, 6]. In general, a phenotype that gives rise to some particular function is generated by a dynamical process that is subject to noise. Hence, phenotypes of isogenic individual organisms are not necessarily identical, and therefore, they form a distribution.

Considering that a biological function is generally a result of such stochastic dynamics, the dynamic process that shapes the function is expected to be robust under such stochasticity; in other words, the phenotype for the function will not be sensitive to noise[7]. However, such robustness is not a general property of dynamics. For example, the complex folding process of a heteropolymer from an arbitrary random sequence might not have such robustness. In this sense, the robustness could be a result of evolution. How can a dynamical process robust to noise be shaped through evolution?

In addition to being robust to noise, a biological system has to remain relatively robust to mutations in the genetic sequence that occur through evolution; the phenotype has to be rather insensitive to changes in the ge-

*Electronic address: ayaka@huku.c.u-tokyo.ac.jp
†Electronic address: hukusima@phys.c.u-tokyo.ac.jp
‡Electronic address: kaneko@complex.c.u-tokyo.ac.jp

netic sequence. Are these two types of robustness correlated? Does noise in the dynamic process affect the evolution of mutational robustness? Indeed, a possible relationship between robustness to noise and robustness to mutation has been discussed [8, 9, 10, 11, 12, 13, 14], following the pioneering study by Waddington[15] on the evolution-development relationship, which is referred to as canalization and genetic assimilation. However, a theoretical understanding of the evolution of robustness is still insufficient.

Consider a dynamic process for shaping a target phenotype. To have robustness to noise in such dynamics, it is ideal to adopt dynamics in which the target phenotype is reached smoothly and globally from a variety of initial configurations and is maintained thereafter. In fact, the existence of such global attraction in the protein folding process was proposed as a consistency principle by Go [16] and as "funnel" landscape by Onuchic et al. [17, 18], while similar global attraction dynamics have been discovered recently in gene regulatory networks [19] and developmental dynamics [20]. In spite of the ubiquity of such funnel-like structures for phenotype dynamics, little is understood about how these structures are shaped by the evolutionary process [1]. We also address this question here and show that it is indeed closely related to the topic of robustness to noise.

In general, it is possible to utilize a complicated model that agrees well with biological reality in order to answer the above-mentioned questions, and this will become necessary in the future. However, at the present level of understanding, in order to understand the concepts in the evolution of robustness, we choose to investigate a rather abstract model that can be made tractable in terms of statistical physics, that is, a system consisting of N Ising spins interacting globally. Each spin can take be either up or down, and each configuration of spins corresponds to a phenotype. The fitness is given by a function of the configuration of some target spins, and this fitness yields a biological function. An equilibrium spin configuration is reached by a certain Hamiltonian that is determined by the interaction between spins. This interaction is given by genes and can change by mutation. By selection according to the fitness function, a Hamiltonian that results in a higher fitness is selected. Indeed, this type of model has been adopted by Saito et al. [1], who utilized it in the study of the evolution of protein folding dynamics, where the spin configuration corresponds to that of residues in a peptide chain, and the folding dynamics are given by decreasing the energy in accordance with the Hamiltonian.

Even though the spin model is abstract, it can account for the basic structures required to study the evolution of genotype and phenotype, i.e., gene → developmental dynamics subject to noise → phenotype → fitness. In comparison with the gene transcription network model utilized in the study of the evolution of robustness [8, 9, 12, 13], the present spin model is computationally efficient in that Monte Carlo simulations and the methods developed in statistical mechanics of spin systems can be applied to answer the above-mentioned general questions on evolution. In fact, we will analyze the evolution of robustness with respect to such a statistical-mechanical framework and define a funnel landscape in terms of frustration, as developed in spin-glass theory [21, 22].

A shorter version of our results has already been published as a letter [23], in which we propose a scenario, based on our numerical simulations, that the ubiquity of funnel-type dynamics observed in biological systems is a consequence of evolutional process under noise beyond a certain level. In this paper, we further study the spin model in particular on the dependence of the result on the number of the total spins N and the target spins responsible for the fitness with extensive numerical simulations. The results suggest that there exists an optimal ratio of the target to the total spins to achieve evolution of robustness over a wide range of temperature. In other words, some degree of redundant spins that do not contribute to the fitness is necessary. We have also carried out statistical-mechanical calculations of the fitness landscape, to provide an interpretation of the relation between the funnel dynamics and robustness to mutations found in our numerical simulation. These findings will stimulate further studies on the understanding on the evolution of robustness from statistical-physics viewpoints.

This paper is organized as follows. In Sec.II, we explain the model setup that captures the essential features of the evolution. In Sec.III, the numerical results for the model are presented. First, we present the dependence of energy and fitness on the temperature and selection process. Then, we show that the evolved Hamiltonians are characterized by the frustration in terms of the statistical physics of spin systems. We classify three phases on the basis of the robustness of the fitness to noise and mutation, and we show that a robust system is realized at an intermediate temperature. The system-size dependence of these results is also discussed. The origin of the robustness is studied in Sec. IV by analytically estimating the fitness landscape by using statistical mechanics. Finally, in Sec.V, the conclusions and prospects for further development are described.

II. MODEL SETUP

We introduce a statistical-mechanical spin model in which the phenotype and genotype are represented by configurations of spin variables S_i and an interaction matrix J_{ij}, respectively, with $i, j = 1, \cdots, N$. The spins S_i and J_{ij} can take one of only two values ± 1, and the interaction matrix is assumed to be symmetric, i.e., $J_{ij} = J_{ji}$. A set of configurations is denoted by $\boldsymbol{S}$ for the phenotype and by $\boldsymbol{J}$ for the genotype. The dynamics of the phenotype are given by a flip-flop update of each spin with an energy function, which is defined by the Hamiltonian for

a given set of genotypes,

$$H(\boldsymbol{S}|\boldsymbol{J}) = -\frac{1}{\sqrt{N}}\sum_{i<j} J_{ij}S_iS_j. \tag{1}$$

We adopt the Glauber dynamics as an update rule, where the N spins are in contact with their own heat bath at temperature T_S. The Glauber dynamics, satisfying the detailed balance conditions, yields an equilibrium distribution for a given $\boldsymbol{J}$:

$$P(\boldsymbol{S}|\boldsymbol{J}, T_S) = \frac{e^{-\beta_S H(\boldsymbol{S}|\boldsymbol{J})}}{Z_S(T_S)}, \tag{2}$$

where $Z_S(T_S) = \mathrm{Tr}_{\boldsymbol{S}} e^{-\beta_S H(\boldsymbol{S}|\boldsymbol{J})}$ and $\beta_S = T_S^{-1}$. After a relaxation process, the phenotype $\boldsymbol{S}$ follows from the equilibrium distribution, and it is not determined uniquely from the genotype $\boldsymbol{J}$; rather, it is distributed, except at zero temperature. The phenotype fluctuation is computed from the Glauber dynamics, and the resulting equilibrium probability distribution. Thus, the degree of fluctuation is characterized by the temperature T_S.

Next, we introduce evolutionary dynamics for the genotype $\boldsymbol{J}$. The genotype is transmitted to the next generation with some variation, while genotypes that produce a phenotype with higher fitness are selected. The time scale for genotypic change is generally much larger than that of the phenotypic dynamics. We assume that the two time scales for the phenotypic expression dynamics and the genotypic evolutionary dynamics are separated, so that the variables $\boldsymbol{S}$ are well equilibrated within the unit time scale of the slow variable $\boldsymbol{J}$. Then, the fitness should be expressed by a function of the phenotype $\boldsymbol{S}$ that is averaged with respect to the distribution. Here, we define the fitness as

$$\Psi(\boldsymbol{J}|T_S) = \Big\langle \prod_{i<j\in\boldsymbol{t}} \delta(S_i - S_j) \Big\rangle \equiv \Big\langle \psi \Big\rangle, \tag{3}$$

where $\langle \cdots \rangle$ denotes the expectation value with respect to the equilibrium probability distribution. The set $\boldsymbol{t}$ denotes a subset of $\boldsymbol{S}$ with size t ; the members of $\boldsymbol{t}$ are termed as target spins. We refer to configurations such that all target spins are aligned in parallel as target configurations, which are assumed to give a requested appropriate function. By a gauge transformation on the target spin and the corresponding elements of $\boldsymbol{J}$, a choice of any other form of spin alignment for the fitness function, instead of the "ferromagnetic" configuration, yields the same result [21]. The fitness can be interpreted as the average frequency of finding the target configurations in equilibrium for a given $\boldsymbol{J}$. It should be noted that in our model, only the target spins contribute explicitly to the fitness and the remaining spins have no direct influence on the fitness and the selection of genes. Hence, the spin configuration for a given fitness has redundancy.

The genotype dynamics are a result of mutations and selection, i.e., changes according to the fitness function following random flip-flops of genes. Hence, for a genetic

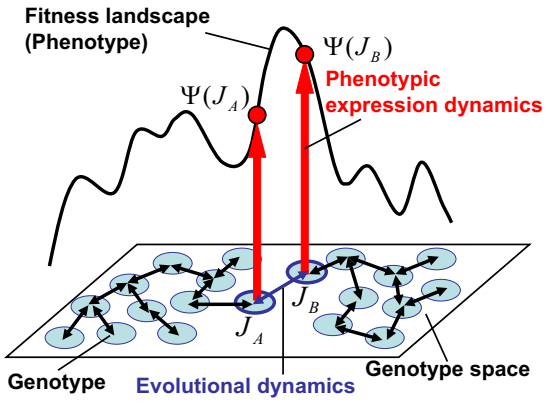

FIG. 1: (color online) A schematic representation of our model. The plane represents the genotype space, where the circles correspond to each configuration of $\boldsymbol{J}$. Their fitness is determined through the phenotypic expression dynamics, which are given by Glauber-type dynamics. The landscape in which the height of each point corresponds to the fitness value is called a fitness landscape. At each generation of genotype evolutional dynamics, genotypes providing higher fitness values are selected under the selection pressure T_J.

dynamics, we once again adopt the Glauber dynamics by using the fitness instead of the Hamiltonian in the phenotype dynamics, where the genotype $\boldsymbol{J}$ is in contact with a heat bath whose temperature T_J is different from T_S. In specific, the dynamics for the genotype are given by a stochastic Markov process with the following stationary distribution:

$$P(\boldsymbol{J}|T_S, T_J) = \frac{e^{\beta_J \Psi(\boldsymbol{J}|T_S)}}{Z_J(T_S, T_J)}, \tag{4}$$

where $Z_J(T_S, T_J) = \mathrm{Tr}_{\boldsymbol{J}} e^{\beta_J \Psi(\boldsymbol{J}|T_S)}$ and $\beta_J = T_J^{-1}$. According to the dynamics, genotypes are selected rather uniformly at high values of the temperature T_J, irrespective of the fitness, whereas at low values of T_J, the genotypes with higher fitness values are preferentially selected. In this sense, the temperature T_J represents the selection pressure among mutated genotypes.

Note that the Glauber dynamics for the genotype $\boldsymbol{J}$ is applied over a much longer time scale than the dynamics for the phenotype $\boldsymbol{S}$; the genotype $\boldsymbol{J}$ changes only during the reproduction of each individual, while the spin dynamics proceed within a developmental time scale to shape the phenotype. Hence, we update J_{ij} after the spins are updated a sufficient number of times for attaining the equilibrium configurations. In actual simulations, a candidate $\boldsymbol{J}'$ for the next generation is set by some flips of a randomly chosen J_{ij} from the current $\boldsymbol{J}$, while the transition probability from $\boldsymbol{J}$ to $\boldsymbol{J}'$ is given by Metropolis rules, $\min(1, \exp(\beta_J(\Psi(\boldsymbol{J}) - \Psi(\boldsymbol{J}'))))$. Fig. 1 shows a schematic explanation of our model.

Our model provides two landscapes: the free energy landscape of spins and the fitness landscape of $\boldsymbol{J}$s. The

free energy landscape is determined by a configuration of $\boldsymbol{J}$ and the phenotypic expression dynamics correspond to the relaxation process on the landscape. The fitness of $\boldsymbol{J}$, $\Psi(\boldsymbol{J})$, is given by the phenotypic expression dynamics on the free energy landscape of spins. We call the landscape of fitness the fitness landscape. The evolutional dynamics of $\boldsymbol{J}$ correspond to a random walk to the top of the fitness landscape; the random walk is generated by noise whose intensity is given by T_J.

A statistical-mechanical spin model similar to ours has been studied for protein evolution [1], where a genetic algorithm is used for genetic dynamics. Our model, which is based on two equilibrium distributions, enables us to conduct this study by using a Markov-chain type simulation, for which a population-based simulation developed by a genetic algorithm is also available. Further, analytical tools developed in statistical mechanics are helpful for gaining a better understanding of the model.

III. RESULTS

A. Fitness and Energy

We have carried out MC simulations of the model discussed above and studied the dependence of the fitness and energy on T_S and T_J. They are given by

$$\Psi(T_S, T_J) = [\Psi(\boldsymbol{J}|T_S)]_J, \quad E(T_S, T_J) = [\langle H(\boldsymbol{S}|\boldsymbol{J})\rangle]_J, \tag{5}$$

respectively, where $[\cdots]_J$ denotes the average with respect to the equilibrium probability distribution, $P(\boldsymbol{J}, T_S, T_J)$. MC sampling with temperature T_S under the Hamiltonian H and the stochastic selection process governed by the fitness are carried out alternately. In our simulations of the spin dynamics, the exchange Monte Carlo simulation (EMC) [24] is introduced to accelerate the relaxation time to equilibrium and obtain the equilibrium spin distribution efficiently. In this section, we concentrate on the analysis of the equilibrium state. Fig. 2 (a) and (b) show the dependence of the fitness and the energy on T_S and T_J, respectively, for $N = 15$ and $t = 3$. For each generation of the genotype dynamics, the fitness and energy are averaged with respect to the equilibrium distribution over 1500 MC steps after discarding the first 1500 MC steps; this number of steps is sufficient for equilibration. The data are averaged over the last 1000 generations. The dependence on the system and target size will be discussed later. For any T_S, the fitness decreases monotonically with T_J, but the rate of decrease is affected significantly by T_S. The fitness for sufficiently low T_S remains at a high level and decreases only slightly with an increase in T_J, while for a medium value of T_S, the fitness gradually decreases to a lower level as a function of T_J, and eventually, for a sufficiently high value of T_S, it never reaches a high level. This result implies that the structure of the fitness landscape depends on T_S, the temperature at which the system has evolved.

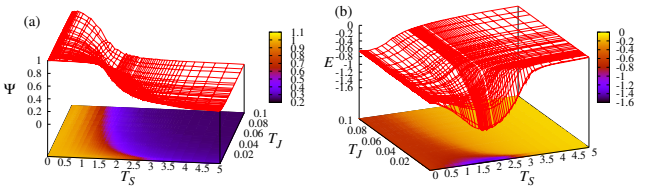

FIG. 2: (color online) The density plots of the fitness $\Psi(\boldsymbol{J})$ and the energy (both are given as Eq. (5)) are shown in (a) and (b), respectively, in the $T_S - T_J$ plane with $N = 15$ and $t = 3$.

The energy function, on the other hand, shows a significant dependence on T_S. While the energy is represented by a monotonically increasing function of T_S, for high T_J, it exhibits non-monotonic behavior for low T_J and has a minimum at $T_S \simeq 2.0$. The configurations that include the target pattern at the energy minimum are obtained around $T_S \simeq 2.0$. This non-monotonicity of the energy corresponds to a negative specific heat in the sense of standard thermodynamics. This is not possible in quenched spin systems with fixed $\boldsymbol{J}$. However, the interactions $\boldsymbol{J}$ depend on the temperature T_S and T_J. It would be convenient to obtain an explicit formula for the derivative of the energy with respect to T_S:

$$\frac{dE(T_S, n)}{dT_S} = \beta_S^2 \Big\{ [\sigma_E^2]_J + \beta_J \mathrm{Cov}_{\boldsymbol{J}} \Big(\langle H \rangle, \mathrm{Cov}_{\boldsymbol{S}}(\psi, H) \Big) \Big\}, \tag{6}$$

where $\mathrm{Cov}_{\boldsymbol{J}}(A, B) = [AB]_J - [A]_J[B]_J$, $\mathrm{Cov}_{\boldsymbol{S}}(A, B) = \langle AB \rangle - \langle A \rangle \langle B \rangle$ and $\sigma_E^2 = \langle H^2 \rangle - \langle H \rangle^2$. The first term of Eq. (6) is the usual specific heat of the random system and it must be positive, and T_S-dependence of $\boldsymbol{J}$ comes from the second term, which can be negative.

The configurations of $\boldsymbol{J}$ giving rise to the highest fitness value generally have a huge redundancy. Using a fluctuation induced by T_S, a specific subset of the configurations of $\boldsymbol{J}$ with lower energy is selected among the redundant configurations at around $T_S \simeq 2.0$.

B. Frustration

In the medium-temperature range, both a lower energy and a higher fitness are achieved. What is the structure in $\boldsymbol{J}$ configurations that helps to achieve this? The statistical physics of spin systems tells us that a decrease in energy implies a decrease in the frustration in spin configurations. By the definition of the Hamiltonian Eq. (1), the possible minimum energy is $-C_2^N/\sqrt{N}$, where C_2^N is the number of the spin pairs. However, if the interaction among the three spins satisfies $J_{ij}J_{jk}J_{ki} < 0$, the energy per spin cannot be minimized to the minimum value $-C_2^N/\sqrt{N}$. Such interactions are said to have frustration[21, 22]. Meanwhile, all the interactions satisfying $J_{ij}J_{jk}J_{ki} > 0$ do not have frustration, and the energy of the spin states attains the minimum

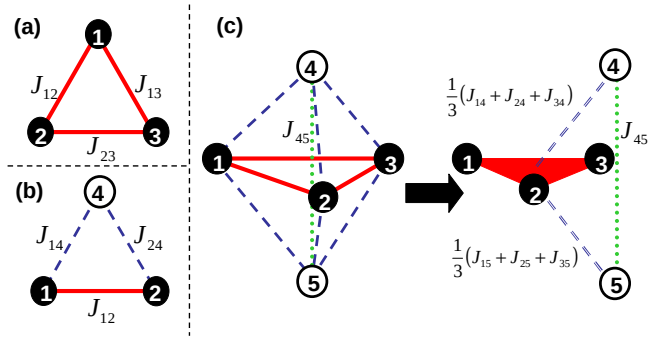

FIG. 3: (color online) Definition of frustration in terms of the parameters Φ_1, Φ_2, and Φ_3 for the case with three target spins ($t = 3$); S_i ($i = 1, 2, 3$) are target spins and the remaining spins are non-targets. The bold lines depict the interactions between target spins ($\in \boldsymbol{J}_{tt}$), the dashed lines depict those between target and non-target spins ($\in \boldsymbol{J}_{to}$), and the dotted lines depict those between non-target spins ($\in \boldsymbol{J}_{oo}$); (a) a triangle consists of three $\boldsymbol{J}_{tt}$ interactions; (b) a triangle consists of one $\boldsymbol{J}_{tt}$ interaction and two $\boldsymbol{J}_{to}$ interactions which are connected to the same non-target spin; (c) a hexahedron that consists of all types of interactions. By summing up the target spins, it is represented as a triangle with a renormalized interaction.

value. However, the energetically favorable spin configuration cannot be uniquely determined only by the condition $J_{ij}J_{jk}J_{ki} > 0$. The spin configurations that have low energy should be the target configurations when both the decrease in energy and the increase in fitness are simultaneously achieved. In our case, the target spins play a distinct role, and therefore, we need to quantify the frustration while distinguishing between target and non-target spins; this is in contrast to the standard spin-glass study.

The interactions are divided into three categories: those between target spins, $\boldsymbol{J}_{tt}$ ($\{J_{ij} \mid i, j \in \boldsymbol{t}\}$), those between target and non-target spins, $\boldsymbol{J}_{to}$ ($\{J_{ij} \mid i \in \boldsymbol{t}, j \in / \boldsymbol{t}\}$), and those between non-target spins, $\boldsymbol{J}_{oo}$ ($\{J_{ij} \mid i, j \notin \boldsymbol{t}\}$). It can be assumed that the frustration of all categories decreases at intermediate T_S. To confirm this, we should define the conditional frustration for each category of spins. Fig. 3(a) shows the minimal configuration consisting of the interactions in $\boldsymbol{J}_{tt}$, (b) shows that consisting of the interactions in $\boldsymbol{J}_{tt}$ and $\boldsymbol{J}_{to}$, and (c) shows that consisting of the interactions in $\boldsymbol{J}_{to}$ and $\boldsymbol{J}_{oo}$.

We first define Φ_1 as the frequency of positive coupling among target spins, i.e.,

$$\Phi_1(T_S, T_J) = \frac{2}{t(t-1)} \Big[\sum_{i<j \in \boldsymbol{t}} J_{ij} \Big]_J. \tag{7}$$

The target configurations are energetically preferred under ferromagnetic coupling, i.e., $\Phi_1 = 1$, for which no frustration exists among the target spins (Fig. 3(a)).

Second, we define Φ_2 as

$$\Phi_2(T_S, T_J) = \frac{2}{t(t-1)(N-t)} \Big[\sum_{i<j \in \boldsymbol{t}} \sum_{k \notin \boldsymbol{t}} J_{ik}J_{jk} \Big]_J, \tag{8}$$

where $t(t-1)(N-t)/2$ is the total number of possible spins: two target spins and one non-target spin. Here, $\Phi_2 = 1$ implies that no frustration exists in the interactions between target and non-target spins, and thus, the target configuration is at an energy minimum even when these interactions are included (Fig. 3(b)).

Lastly, as a measure of the frustration among non-target spins, Φ_3 is defined as

$$\Phi_3(T_S, T_J) = \frac{1}{C_2^{N-t}} \Big[\sum_{k<l \notin \boldsymbol{t}} \Big(\frac{1}{t} \sum_{i \in \boldsymbol{t}} J_{ik} \Big) J_{kl} \Big(\frac{1}{t} \sum_{j \in \boldsymbol{t}} J_{jl} \Big) \Big]_J, \tag{9}$$

where C_2^{N-t} is the total number of possible pairs of non-target spins. Fig. 3(c) helps to comprehend the definition of Φ_3. Each non-target spin interacts with all the target spins, and it has t interactions that are categorized into $\boldsymbol{J}_{to}$, e.g., J_{14}, J_{24} and J_{34}. By summing up all the interactions, the frustration is computed as $\frac{1}{9}(\sum_{i=1}^{3} J_{i4})J_{45}(\sum_{j=1}^{3} J_{5j})$. By considering all the possible non-target spins instead of sites 4 and 5, Φ_3 is defined as Eq. (9). If Φ_3 is equal to 1, the frustration is not introduced by the interactions between non-target spins; in other words, there is no frustration globally. Hence, the system with $\Phi_1 = \Phi_2 = \Phi_3 = 1$ is in the Mattis state [25], which can be transformed to ferromagnetic interaction by a gauge transformation.

For the interaction $\boldsymbol{J}$, with evolution under an environment with temperature T_S, we have computed Φ_1, Φ_2, and Φ_3 by performing MC simulations. In Fig. 4, we present contour maps of (a) $\Phi_1(T_S, T_J)$, (b) $\Phi_2(T_S, T_J)$, and (c) $\Phi_3(T_S, T_J)$ in the $T_S - T_J$ plane. At sufficiently low T_J, the frustration parameters attain the maximum value 1 at the intermediate T_S, where the frustrations are extensively eliminated, while they remain finite at low T_S and high T_S. We define the intermediate temperature region as $T_S^{c1} < T_S < T_S^{c2}$, where the frustration parameter Φ_2 equals 1. These temperatures depend on T_J, and we express them as $T_S^{c1}(T_J)$ and $T_S^{c2}(T_J)$. We see that for low T_J ($\lesssim 0.05$), the phase diagram is split into three phases. The first one is frustrated and adapted phases for $T_S < T_S^{c1}(T_J)$. For $T_S < T_S^{c1}(T_J)$, all Φ_i ($i = 1, 2, 3$) are less than unity, and hence, the frustration remains for target and non-target spins.

For $T_S \geq T_S^{c1}(T_J)$, Φ_1 equals 1, so that a target configuration is embedded as an energetically favorable state (Fig. 4(a)). For a finite system with finite T_J, Φ_j cannot be exactly 1. However, as long as T_J is low, the deviation of Φ_j from 1 at the intermediate temperature is negligible. In contrast to Φ_1, the sum of the J_{ij} in $\boldsymbol{J}_{to}$ and $\boldsymbol{J}_{oo}$ fluctuates around 0 at any T_S. Fig. 5 shows the averages $\overline{\boldsymbol{J}_{to}} = [\sum_{i \in \boldsymbol{t}, j \in \boldsymbol{o}} J_{ij}]_J$ and $\overline{\boldsymbol{J}_{oo}} = [\sum_{i \in \boldsymbol{o}, j \in \boldsymbol{o}} J_{ij}]_J$ of the summation of J_{ij} in $\boldsymbol{J}_{to}$ and $\boldsymbol{J}_{oo}$, respectively, at a low T_J (0.5×10^{-3}). As shown in Fig. 5, $\overline{\boldsymbol{J}_{to}}$ and $\overline{\boldsymbol{J}_{oo}}$

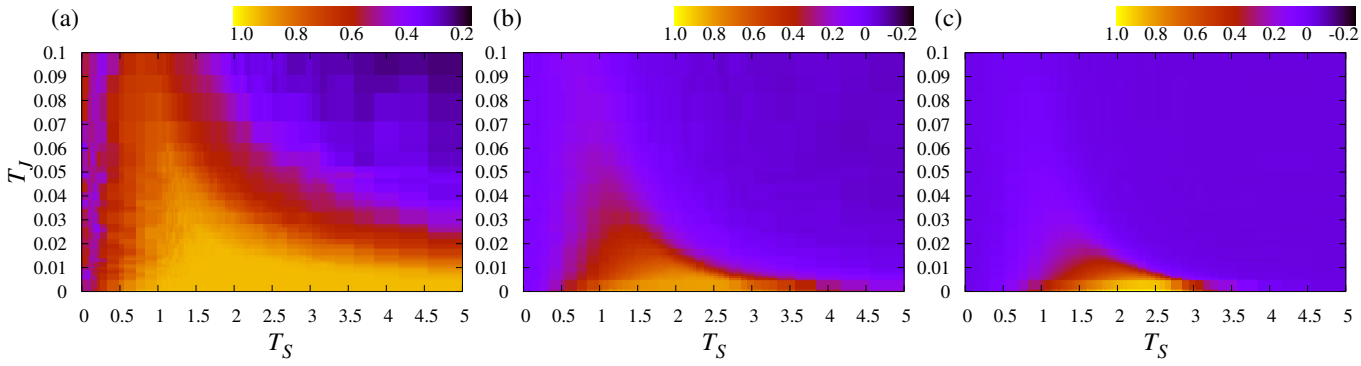

FIG. 4: (color online) Density plot of local frustrations: (a) Φ_1, (b) Φ_2, and (c) Φ_3 in the $T_S - T_J$ plane. The data are computed by averaging over 150 genotypes $\boldsymbol{J}$ evolved at given temperatures T_S and T_J with $N = 15$ and $t = 3$.

do not deviate from 0 at any T_S. This implies that no specific patterns are embedded in the spin configuration apart from the target spins.

For $T_S^{c1}(T_J) \leq T_S \leq T_S^{c2}(T_J)$, Φ_2 is also equal to 1, implying that the frustration among spins is not introduced via interactions with a non-target spin (Fig. 4(b)). In this temperature range, Φ_3 is not always equal to 1, except for $T_S \sim 2.0$, where the Mattis state arises (Fig. 4(c)). When $\Phi_2 = 1$ and $\Phi_3 \neq 1$, the frustration is not completely eliminated from the non-target spin interactions $\boldsymbol{J}_{oo}$; this is in contrast to the Mattis state. We call such a $\boldsymbol{J}$ configuration "local Mattis state," and it is characterized by $\Phi_1 = \Phi_2 = 1$ but $\Phi_3 \neq 1$. This implies that the interactions $\boldsymbol{J}$ have no frustration around the target spins, but there is some frustration between non-target spin interactions. The interactions $\boldsymbol{J}$ required to form such a local Mattis state are obtained as a consequence of the evolution around $T_S^{c1}(T_J) \leq T_S \leq T_S^{c2}(T_J)$ for low T_J, where both the fitted target configuration and lower energy are achieved. The T_S range in which the local Mattis state is stabilized becomes narrower with an increase in T_J. The phase diagram of the model is shown in Fig. 6.

For $T_S > T_S^{c2}(T_S)$, the frustration parameter Φ_2 is less than 1, and consequently, the frustration remains, and the fitness Ψ starts to decrease and the energy increases. Thus, neither adaptation nor energy minimization is achieved. The parameters Φ_is should converge to 0 as $T_S \to \infty$, because for random $\boldsymbol{J}$, the numbers of frustrated and non-frustrated loops are equal. In fact, at $T_S \sim 5.0$, Φ_1 also starts to decrease.

C. Relaxation dynamics

Thus far, we have computed the fitness in the equilibrium state by using EMC for accelerating the relaxation dynamics of spins. Under standard Glauber dynamics, the process may require much longer time steps. Here, we discuss the relaxation dynamics of spins for adapted interactions $\boldsymbol{J}$ that are obtained from evolution under

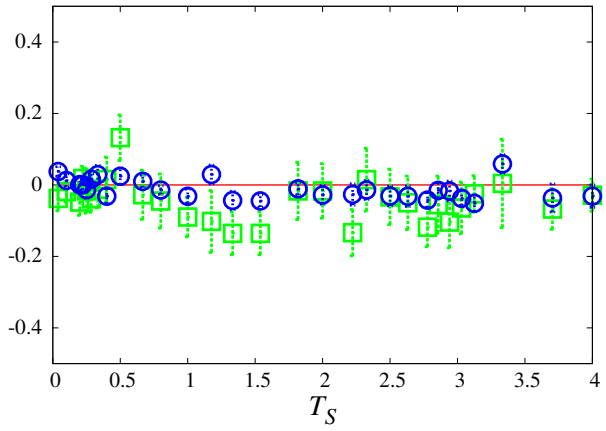

FIG. 5: (color online) T_S-dependence of the averages $\overline{\boldsymbol{J}_{to}}$ (□) and $\overline{\boldsymbol{J}_{oo}}$ (◯) for a fixed $T_J = 0.5 \times 10^{-3}$.

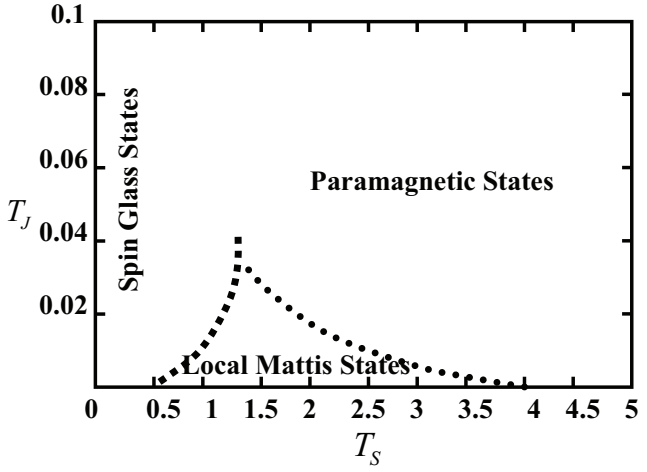

FIG. 6: Phase diagram of the evolved $\boldsymbol{J}$s at $N = 15$ and $t = 3$. Three types of evolved $\boldsymbol{J}$ are defined on the basis of the value of the fitness and Φ_2. Their properties are summarized below in Table I.

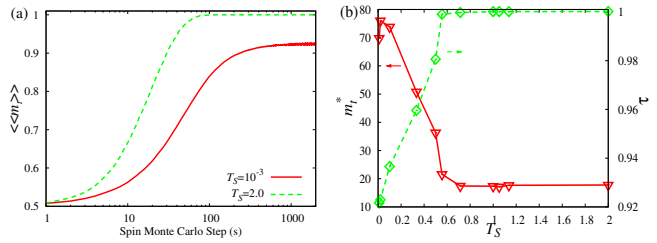

FIG. 7: (color online) (a) Relaxation dynamics of the averaged magnetization of target spins, $\langle\langle m_t\rangle\rangle$, averaged over the adapted interactions $\boldsymbol{J}$ for $T_S = 10^{-3}$ (solid curve) and $T_S = 2.0$ (dashed curve). The magnetization $\langle\langle m_t\rangle\rangle$ is evaluated by calculating the average over 30 initial conditions for each $\boldsymbol{J}$ and 1000 different samples of $\boldsymbol{J}$ that are drawn from $P(\boldsymbol{J}, T_S, T_J = 10^{-3})$.
(b)The T_S-dependence of the estimated convergent value of $\langle\langle m_t\rangle\rangle$, m_t^* on the right axis and the relaxation time τ on the left. The relaxation time is estimated from the time constant of an exponential decay of $\langle\langle m_t\rangle\rangle$.

the condition of given T_S and T_J. The target magnetization $m_t = |\frac{1}{t}\sum_{i\in\boldsymbol{t}} S_i|$ is computed as a function of time. Note that we do not use EMC here; rather, we adopt standard MC to directly observe the energy landscape of $\boldsymbol{J}$ adapted through evolution. We calculate the average of m_t over $\boldsymbol{J}$ drawn from an equilibrium distribution $P(\boldsymbol{J}, T_S, T_J)$ at T_S and T_J. Fig. 7 (a) shows the relaxation dynamics of $\langle\langle m_t\rangle\rangle$ for $T_S = 10^{-3} (\leq T_S^{c1})$ and $T_S = 2.0 (T_S^{c1} \leq T_S \leq T_S^{c2})$, where $\langle\langle\cdots\rangle\rangle$ denotes the average over the initial conditions randomly chosen and over interactions $\boldsymbol{J}$, according to $P(\boldsymbol{J}, T_S, T_J)$. In the simulation, we choose a sufficient low $T_J (= 10^{-3})$ so that the obtained interactions have high fitness values. A common working temperature $T_S' (= 10^{-5})$ for relaxation is also chosen to be very low in order to examine the T_S-dependence of the adapted interactions $\boldsymbol{J}$ at T_S and T_J. By performing this simulation, the landscape properties of the typical $\boldsymbol{J}$ adapted at T_S are clearly determined from a dynamical viewpoint. As shown in Fig. 7, the relaxation process of $\langle\langle m_t\rangle\rangle$ for low T_S is much slower even when the working temperatures T_S' are the same. Furthermore, the relaxation process converges to a value m_t^* that is less than 1 and remains at that value for a long time. The deviation of $\langle\langle m_t\rangle\rangle$ from 1 gives the fraction of the initial conditions that fails to reach the target within this time span, because each m_t for $t = 3$ is either 1 or 1/3 depending on whether the target configuration is reached. Indeed, the relaxation dynamics are strongly dependent on the initial conditions. For some initial conditions, the spins are trapped at a local minimum, and as a result, the target configuration is not realized over a long time span. After a much longer time span, $\langle\langle m_t\rangle\rangle$ approaches 1, the equilibrium value, when the spins are updated under the temperature T_S, i.e., the temperature adopted for evolution. Such dependence on initial conditions is not observed for $\langle\langle m_t\rangle\rangle$ when $T_S > T_S^{c1}$, where $\langle\langle m_t\rangle\rangle$ approaches 1 rather quickly.

From an estimate of the convergent value of the target magnetization, m_t^*, within a given time scale, we obtain the relaxation time τ by fitting the estimates to the function $\langle\langle m_t\rangle\rangle(s) = m_t^* + c\exp(-s/\tau)$, where s is the Monte Carlo step of the spin dynamics. The parameters m_t^* and τ are plotted against T_S in Fig. 7(b), which shows that τ starts to increase and m_t^* decreases from 1 as T_S decreases below T_S^{c1}. These results imply that the interactions $\boldsymbol{J}$ whose energy landscapes are rugged, similar to the energy landscape of a spin-glass phase, are dominant for $T_S \leq T_S^{c1}$, whereas those with a smooth landscape around the target are dominant for $T_S^{c1} \leq T_S \leq T_S^{c2}$. The latter can be interpreted as a type of funnel landscape. Our result supports the occurrence of transitions from the spin-glass phase to the funnel phase at T_S^{c1}caused by thermal fluctuation. Note that the evolutional formation of funnel from rugged landscapes was also observed by Saito et al. in the evolution simulations of spin systems for protein folding [1].

D. Robustness to mutation

We now examine the mutational robustness of the evolved genotypes in detail. The robustness to mutations corresponds to the stability of the fitness of $\boldsymbol{J}$ with respect to changes in the $\boldsymbol{J}$ configuration. From the genotypes $\boldsymbol{J}$ that are generated by $P(\boldsymbol{J}|T_S, T_J)$, mutations are imposed by flipping the sign of a certain fraction of randomly chosen matrix elements in $\boldsymbol{J}$. The value of the fraction corresponds to the mutation rate μ. We evaluate the fitness of the mutated $\boldsymbol{J}'(\boldsymbol{J}, \mu)$ at T_S', i.e.,

$$[\Psi(\boldsymbol{J}'(\boldsymbol{J},\mu)|T_S')]_{J(T_S,T_J)} = \mathop{\mathrm{Tr}}_{\boldsymbol{J}} P(\boldsymbol{J}|T_S,T_J)\frac{\mathop{\mathrm{Tr}}_{\boldsymbol{S}}\psi e^{-\beta_S' H(\boldsymbol{S}|\boldsymbol{J}'(\boldsymbol{J},\mu))}}{Z(T_S', \boldsymbol{J}'(\boldsymbol{J},\mu))}, \quad (10)$$

where $\beta_S' = 1/T_S'$ and $Z(T_S', \boldsymbol{J}') = \mathop{\mathrm{Tr}}_{\boldsymbol{S}} e^{(-\beta_S' H(\boldsymbol{S}|\boldsymbol{J}'))}$. The bracket $[\cdots]_{J(T_S,T_J)}$ is almost identical to that denoted by $[\cdots]_J$ defined above; however, the additional subscript (T_S, T_J) indicates the temperatures at which the genotype $\boldsymbol{J}$ evolves. If $\mu = 0$ and $T_S' = T_S$, $[\cdots]_{J(T_S,T_J)}$ is equal to the usual fitness defined in Eq. (5); $[\Psi(\boldsymbol{J}'(\mu = 0)|T_S' = T_S)]_{J(T_S,T_J)} = \Psi(T_S, T_J)$. In order to distinguish the mutational robustness from thermal noise, we set $T_J = 0.5 \times 10^{-3}$ to ensure that the fitness value with $\mu = 0$ is 1 and the working temperature is $T_S' = 10^{-5}$. The fitness averaged over 150 samples of mutated $\boldsymbol{J}'$ is plotted against the mutation rate μ in Fig. 4 in [23] for $T_S = 10^{-4}$ and $T_S = 2.0$. For low T_S, the fitness of mutated $\boldsymbol{J}$ exhibits a rapid decrease with an increase in the mutation rate, but when T_S is between T_S^{c1} and T_S^{c2}, the fitness does not decrease until the mutation rate reaches a specific value. We define $\mu_c(T_S)$ as a threshold mutation rate beyond which the fitness $[\Psi(\boldsymbol{J}'(\boldsymbol{J},\mu)|T_S')]_{J(T_S,T_J)}$ drops below 1. The value μ_c has a plateau at $T_S^{c1} \leq T_S \leq T_S^{c2}$. The range of tempera-

Phase	Adaptation	Frustration	Landscape	Robustness
SG	Adapted	Frustrated	Rugged	Not robust
LMS	Adapted	Not frustrated	Funnel	Robust
PM	Not adapted	Frustrated	Rugged	Not robust

TABLE I: Three types of $\boldsymbol{J}$—spin-glass (SG), local Mattis state (LMS), and paramagnetic state (PM)—are defined by the value of the local frustration parameter Φ_2 (Fig. 6). The adaptation and frustration of $\boldsymbol{J}$s have been studied in the previous sections, and their landscape and robustness at a working temperature T_S' have been studied in this section. We have adopted a very low temperature T_S' to reveal the energy and fitness landscape given by the interaction matrix $\boldsymbol{J}$ evolved at each temperature T_S. The $\boldsymbol{J}$s evolved at $T_S < T_S^{c1}$ (these $\boldsymbol{J}$s belong to the SG phase) have high fitness values and frustrations. Their adaptation is not robust to noise and mutation because of their rugged landscapes. The $\boldsymbol{J}$s evolved at $T_S^{c1} < T_S < T_S^{c2}$, (these $\boldsymbol{J}$s belong to the LMS phase) have high fitness values and less frustrations. They are robust to noise and mutation, and their landscapes give funnel-type dynamics for spins. The fitness values of the $\boldsymbol{J}$s evolved at $T_S > T_S^{c2}$ (these $\boldsymbol{J}$s belong to the PM phase) cannot be high, and the $\boldsymbol{J}$s have frustrations.

tures that result in mutational robustness as evolution proceeds agrees with the range giving rise to the local Mattis state, where Φ_2 is unity. In other words, mutational robustness is realized for a set of genotypes with no frustration around the target spins. The evolution to a mutationally robust genotype $\boldsymbol{J}$ is possible only when the phenotype dynamics are subjected to noise in the range $T_S^{c1} \leq T_S \leq T_S^{c2}$.

To summarize, there are three phases, local Mattis, spin glass and paramagnetic disorder ones, for the evolved $\boldsymbol{J}$, as shown in Fig. 6. Each phase defined by the frustration parameter Φ_2 has a distinct characteristic feature in relaxation dynamics and some robustness, that are summarized in Table I.

E. Size dependence and the existence of an optimal target size to obtain LMS

To check the generality of the transition to the local Mattis state as well as the mutational robustness, we examine the model for three system sizes, $N = 15$, 20, and 30; the ratio $t/N = 0.2$. In Fig. 8, we compare the T_S-dependence of (a) the fitness, (b) the energy, and (c) one of the frustration parameters Φ_2 at $N = 15$ ($t = 3$), $N = 20$ ($t = 4$) and $N = 30$ ($t = 6$) at a fixed $T_J = 0.5 \times 10^{-3}$. As shown in Fig. 8, the behavior of these quantities is qualitatively similar, and by rescaling the temperature T_S by a factor $\sqrt{N}$, the fitness, energy, and Φ_2 lines merge into a single line until the PM phase appears (insets of Fig. 8). The plateau $\Phi_2 = 1$ exists at all system sizes we have studied, and we show that the rescaled temperature $T_S^{c1}(N)/\sqrt{N}$ fit together.

The rescaling factor is $\sqrt{N}$ since the order of energy changes from $O(N)$ to $O(N^{3/2})$ by the evolution at the intermediate T_S because of the existence of the local Mattis states. We have defined the Hamiltonian Eq. (1) with the normalization coefficient $1/\sqrt{N}$; in this definition, we consider the average over $\boldsymbol{J}$ that would be drawn from an i.i.d set of $\boldsymbol{J}$s. However, at intermediate T_S, the distribution of $\boldsymbol{J}$ deviates from the uniform distribution, and local Mattis states appear with high probability. As a result, the order of energy changes, and the temperature $T_S^{c1}(N)$ is proportional to $\sqrt{N}$.

Next, we change the number of target spins t while fixing N at 30. Fig. 9 shows (a) the fitness, (b) the energy, and (c) the frustration parameter Φ_2, at $N = 30$ and $3 \leq t \leq 15$. The threshold temperature T_S at which the fitness value decreases rapidly increases as t increases from a small value to 9 and decreases for larger t. The T_S-dependence of the energy shows non-monotonic behavior, indicating the existence of the local Mattis states for all the values of t studied here. Similar to the threshold temperature obtained from the fitness value, the temperature at which the energy takes a minimum value shows non-monotonic behavior as t increases. Further, the range of the plateau with $\Phi_2 = 1$ is maximized at $t = 7$ and 8. Eventually, the plateau vanishes at $t = 12$ at least. The PM phase, where the adaptation is not acquired, extends toward the lower temperature region with an increase in t; the LMS phase becomes narrower as t increases. Fig. 9 shows that the LMS exists only up to $t \lesssim 0.4N$, and for $t > 0.4N$, a direct transition from the SG to PM phase occurs with an increase in T_S. The LMS region in which robustness and high fitness can be achieved is largest at $t \sim N/4$. These findings indicate that there exists an upper limit of t over which the local Mattis states cannot be obtained and that there exists a suitable value of t for stabilizing the local Mattis states for a wide range of T_S. The above simulation results are summarized in Fig. 10, where the region of local Mattis state is displayed in the $t/N - T_S$ plane, by fixing N and T_J at 30 and 0.5×10^{-3}, respectively.

Finally, we remark on the equilibration time of the evolutional dynamics of $\boldsymbol{J}$ to achieve an adapted state, by varying the value of t. At a fixed T_J, the time scale increases with the increase of t, even for a fixed size N. Particularly, this increase is significant in the SG phase, while it is moderate in LMS phase. This might remind us of slow relaxation of the phenotype dynamics discussed in the previous section.

IV. FITNESS LANDSCAPE

We now explain why mutational robustness is realized only in the intermediate range of temperatures T_S. We do so by performing a statistical-mechanical calculation of the number of fitted states in order to obtain a sketch of the fitness landscape for $\boldsymbol{J}$. We estimate the degeneracy of the states with highest fitness as $T_S \to 0$. Let $W(E_n(\boldsymbol{J}))$ and $W_\Psi(E_n(\boldsymbol{J}))$ be the number of n-th excited states and the number of the target configurations

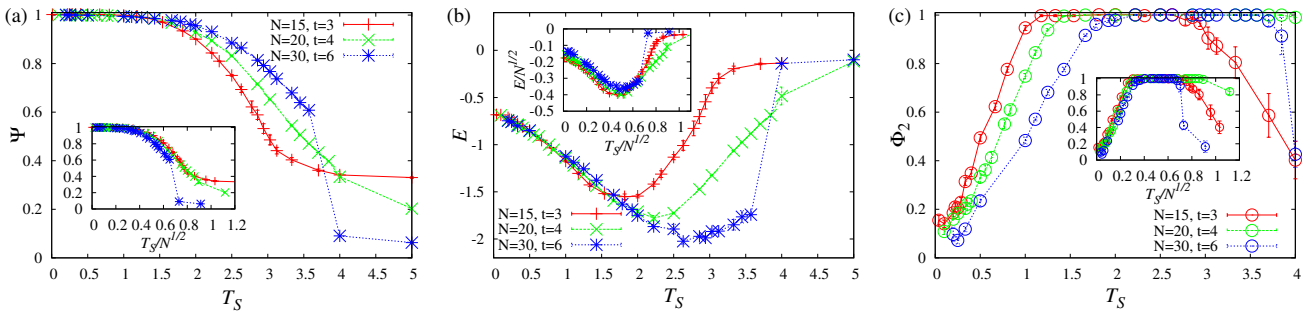

FIG. 8: (color online) The T_S-dependence of (a) the fitness, (b) the energy, and (c) the frustration parameter Φ_2 at $N = 15,\ t = 3,\ N = 20,\ t = 4$ and $N = 30, t = 6$. The temperature T_J is fixed at 0.5×10^{-3}. The inset of each figure shows the dependence of each quantity on the temperature rescaled by $\sqrt{N}$. The vertical axis of the inset of (b) is also rescaled by $\sqrt{N}$.

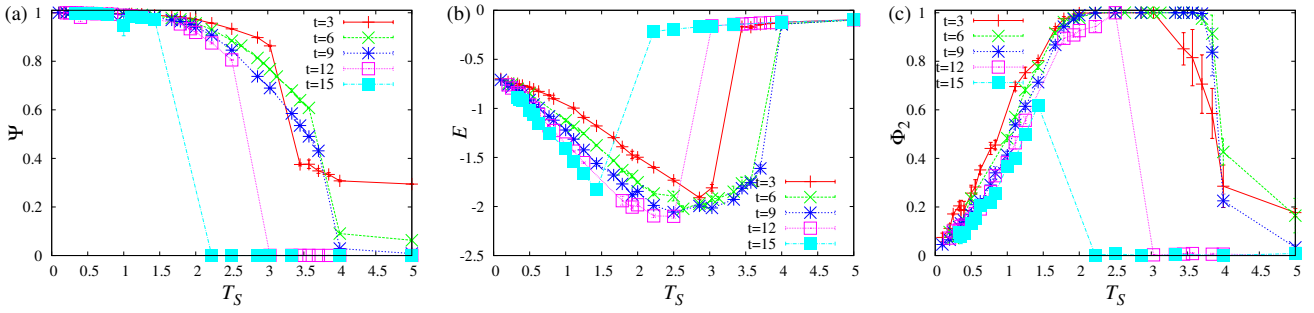

FIG. 9: (color online) The T_S-dependence of (a) the fitness, (b) the energy, and (c) Φ_2 at $N = 30$ and $t = 3, 6, 9, 12$, and 15. The temperature T_J is fixed at 0.5×10^{-3}. The convergence value of the fitness (a) at $T_S \to \infty$ is 2^{-t+1}.

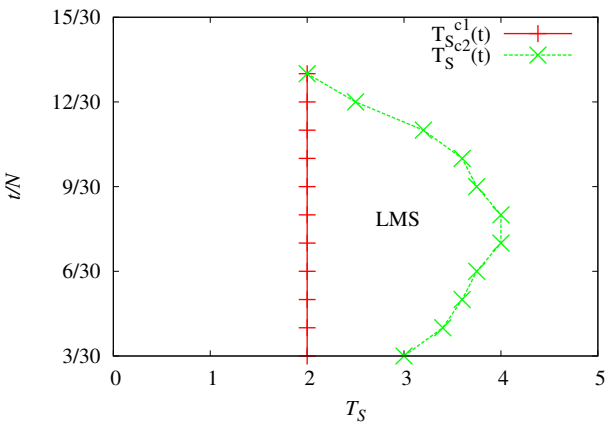

FIG. 10: (color online) $t/N - T_S$ diagram of the domain where the local Mattis states are found at $N = 30$ and a fixed $T_J = 0.5 \times 10^{-3}$. The points on the boundary of the domain are estimated by the values of T_S at which Φ_2 drops from 1, as in Fig. 9(c), and the line is guide to eye.

of n-th excited states for a given $\boldsymbol{J}$. Then, the fitness in the limit as $T_S \to 0$ is given by

$$\Psi(\boldsymbol{J}, T_S) = \frac{\sum_n W_\Psi(E_n(\boldsymbol{J}))e^{-\beta_S E_n(\boldsymbol{J})}}{\sum_n W(E_n(\boldsymbol{J}))e^{-\beta_S E_n(\boldsymbol{J})}} \equiv \frac{Z_\Psi(\boldsymbol{J}, T_S)}{Z(\boldsymbol{J}, T_S)} \longrightarrow \frac{W_\Psi(E_0(\boldsymbol{J}))}{W(E_0(\boldsymbol{J}))}. \quad (11)$$

Accordingly, as $T_S \to 0$, the highest fitness, i.e., unity, is achieved if and only if $\boldsymbol{J}$ satisfies the condition $W_\Psi(E_0(\boldsymbol{J})) = W(E_0(\boldsymbol{J}))$. In fact, a large number of $\boldsymbol{J}$ satisfy this condition besides the local Mattis state. For example, let us consider a Mattis state $\boldsymbol{J}$ with no frustration at all and introduce several changes in the sign of the bonds between target spins $\boldsymbol{J}_{tt}$, target and non-target spins $\boldsymbol{J}_{to}$, and non-target spins $\boldsymbol{J}_{oo}$. This procedure, if applied only to the bond flips to $\boldsymbol{J}_{oo}$, produces the local Mattis states.

First, we calculate the fitness of the Mattis states ($\Phi_1 = \Phi_2 = \Phi_3 = 1$), and we consider representative examples of the local Mattis states ($\Phi_1 = \Phi_2 = 1, \Phi_3 \neq 1$) and target-frustrated states that have frustration among the target spins and hence $\Phi_1 \neq 1$. Their fitness is given as the ratio of the "conditioned partition function" $Z_\Psi(\boldsymbol{J})$ and the partition function $Z(\boldsymbol{J})$, as given by Eq. (11). The location of the frustrations does not influence the partition function, but it influences $Z_\Psi(\boldsymbol{J})$. To determine their fitness function, we should derive the

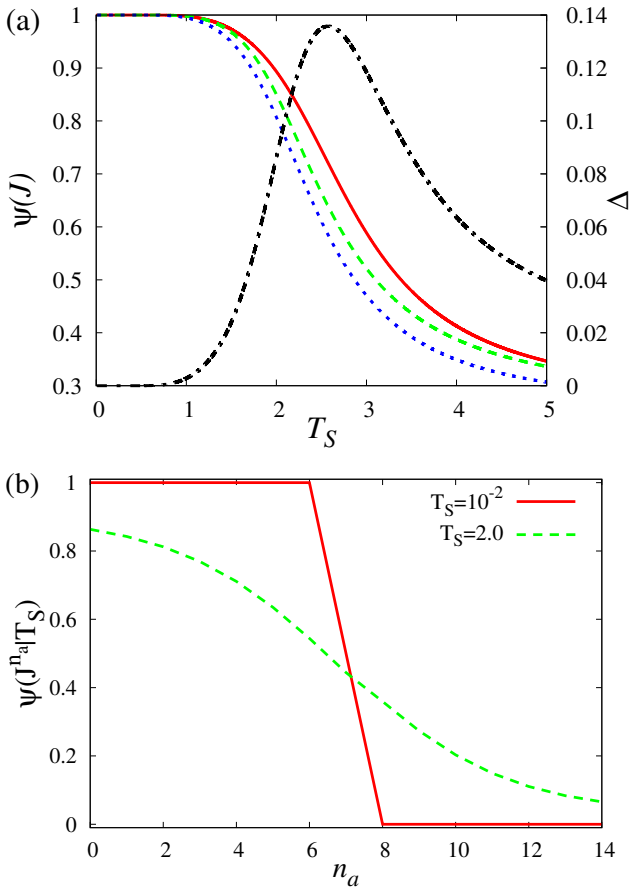

FIG. 11: (color online) (a) The T_S-dependence of the fitness of the Mattis state (bold line, $\Phi_1 = \Phi_2 = \Phi_3 = 1$), one of the local Mattis states (dashed line, $\Phi_1 = \Phi_2 = 1, \Phi_3 = 0.89$), and one of the target-frustrated states (dotted line, $\Phi_1 = 0.333, \Phi_2 = 1, \Phi_3 = 0.89$), plotted on the left axis. The dashed-dotted line represents the difference Δ between the fitness of the Mattis state and that of the target-frustrated state; Δ is plotted on the right.
(b) The fitness of $\tilde{\boldsymbol{J}}^{n_a}$ adapted at $T_S = 10^{-2}$ (solid line) and $T_S = 2.0$ (dashed line) with $N = 15$ and $t = 3$, plotted as a function of the number n_a of flipped bonds of $\tilde{\boldsymbol{J}}^{n_a}$. At low T_S, the fitness decreases rapidly for $n_a = 7$ and decreases gradually at intermediate T_S.

partition functions $Z(\boldsymbol{J})$ and $Z_\Psi(\boldsymbol{J})$ for three types of $\boldsymbol{J}$.

At first, the partition function of $\boldsymbol{J}$ with x frustrated interactions, $Z(x, T_S)$ is given as

$$Z(x, T_S) = 2c(x) \sum_{i=0}^{x} C_i^x f(i, N), \tag{12}$$

where

$$f(i, N) = \left(4e^{\frac{2\beta_S}{\sqrt{N}}} \sinh(\frac{2\beta_S}{\sqrt{N}})\right)^i \sum_{n=i}^{N-i} e^{-\frac{\beta_S}{\sqrt{N}}(N-n)n} C_{n-i}^{N-2i}, \tag{13}$$

and $c(x) = e^{\beta_S(-2x+C_2^N)/\sqrt{N}}$. This expression is valid for the case where there is at most one flipped bond at each site. Therefore, x should be less than $N/2$, and the expression Eq. (12) is efficient, independent of how to assign the value x for $\boldsymbol{J}_{tt}$, $\boldsymbol{J}_{to}$, and $\boldsymbol{J}_{oo}$.

Next, the conditioned partition function of the local Mattis state with x frustrated $\boldsymbol{J}_{oo}$ interactions (denoted as $Z_\Psi^{LMS}(x, T_S)$) is given as

$$Z_\Psi^{LMS}(x, T_S) = 2c(x) \sum_{i=0}^{x} C_i^x f(i, N-t). \tag{14}$$

This expression is valid for the case where there is at most one flipped bond at each site. Accordingly, x should be less than $(N-t)/2$. The fitness of the local Mattis states with x frustrated interactions in $\boldsymbol{J}_{oo}$ is given as $Z_\Psi^{LMS}(x, T_S)/Z(x, T_S)$. Furthermore, the conditioned partition function of the target frustrated states that have y frustrations in $\boldsymbol{J}_{tt}$ (implying $\Phi_1 \neq 1$) and x frustrations in $\boldsymbol{J}_{oo}$ (denoted as $Z^{TF}(x, y, T_S)$), is derived from Eq. (13). When a bond between a pair of target spins is flipped from the Mattis state and frustration is generated among the target spins, the energy of the target configuration increases by $2/\sqrt{N}$; therefore,

$$Z_\Psi(x, y, T_S) = Z_\Psi^{LMS}(x, T_S) \times e^{-\frac{2\beta_S}{\sqrt{N}}y}, \tag{15}$$

where we again consider that each target spin is connected at most one flipped bond and $y = 1, \cdots, t/2$. The partition functions of the target-frustrated states with x frustrated $\boldsymbol{J}_{oo}$ bonds and y frustrated $\boldsymbol{J}_{tt}$ bonds are given by $Z(x + y, T_S)$, and their fitnesses are given by $Z_\Psi^{TF}(x, y, T_S)/Z(x + y, T_S)$.

Fig. 11 shows the T_S-dependence of the typical fitness values for Mattis, local Mattis, and target-frustrated states. The difference between the fitness of the target-frustrated state and that of the Mattis state, which is denoted as Δ, is also plotted as a function of T_S. The value of fitness always approaches unity as $T_S \to 0$, whereas such degeneracy is split by an increase in T_S. From the difference in fitness between the Mattis and the target-frustrated states, Δ, the ratio of the probabilistic weight between these states is obtained as $\exp(\beta_J \Delta)$. This suggests that fewer frustrated $\boldsymbol{J}$ states around the target, i.e., the local Mattis states, are preferentially selected only at the intermediate temperature.

Next, we introduce frustration into $\boldsymbol{J}_{to}$ and denote the constructed $\boldsymbol{J}$ as $\tilde{\boldsymbol{J}}_t^{n_a}$, where the superscript n_a represents the number of altered bonds and the subscript t represents the condition in which the altered bonds exist between the target and non-target spins, i.e., $\boldsymbol{J}_{to}$. Therefore, the frustration parameter Φ_2 for $\boldsymbol{J}_{to}$ of the state $\tilde{J}_t^{n_a}$ does not equal 1. The state $\tilde{J}_t^0$ is simply the original Mattis state, which has the highest fitness, whereas for $n_a = N - 1$, direct computation shows that the fitness is the least. Again, from a straightforward calculation, it can be shown that the fitness of $\tilde{J}_t^{n_a}$ remains to be the highest fitness up to $n_a \leq N/2$. Hence, there is a region in the $\boldsymbol{J}$-state space connected by a single point mutation (change of sign in a single element in $\boldsymbol{J}$) in which the fitness retains its highest value. We refer to this region as

the neutral space[2, 8, 26], in the sense that a mutation within the region is neutral. Note that in addition to this construction, there is more degeneracy among the fittest $\boldsymbol{J}$, as shown in Fig. 11(a).

We have computed how the fitness decreases as $\boldsymbol{J}$ is changed to leave the neutral space, for $T_S \to 0$. In Fig. 11(b), the fitness of $\tilde{J}_t^{n_a}$ is plotted as a function of the number of altered bonds n_a. By just a single point mutation, the fitness decreases suddenly to its lowest value at some n_a. This implies the existence of a clear edge in the neutral space. The genotype located at the edge of the neutral space is not robust to mutation. It is obvious that the Mattis state, i.e., the genotype located at the center of the neutral space, is robust to mutation. However, since the fitness of genotypes is constant throughout the neutral space, both the robust genotypes at the center of the neutral space and the non-robust genotypes at the edge are selected with equal weight. Then, there is no selection pressure to eliminate genotypes that are at the edge of the neutral space. Although complete degeneracy holds only for $T_S \to 0$, the above argument is valid for sufficiently low temperatures, and therefore, the robustness to mutation cannot be expected to exist in the spin-glass phase at low T_S. This is also related to the fact that long time scale for equilibration in the evolutional dynamics is required in the spin-glass phase. As shown in Fig. 11(b), the fitness value around an adapted $\boldsymbol{J}$ decreases abruptly against the mutation. The number of $\boldsymbol{J}$ configurations on the plateau with $\Psi = 1$ that appear at low T_S is roughly estimated as 2^{N-t} because of the gauge transformations for $N - t$ sites, while the total number of possible $\boldsymbol{J}$ configurations is 2^{N^2}. The ratio $2^{N-t}/2^{N^2}$ is strongly suppressed as N increases. This implies that the evolution of $\boldsymbol{J}$ hardly finds the plateau by the local update.

In contrast, at intermediate T_S, the fitness landscape is not neutral. For example, the fitness of $\tilde{\boldsymbol{J}}_t^{n_a}$ gradually decreases from its highest value with increasing n_a, as shown in Fig. 11(b). There is selection pressure toward the genotype with $n_a = 0$. The genotypes with larger n_a have both lower robustness to mutation and lower fitness, but less of such genotypes are selected. Hence, the evolution toward higher fitness also induces robustness to mutation, as a result of the correlation between fitness and robustness.

On the basis of the above argument, schematic representations of the fitness landscape at low and intermediate T_S, together with the distribution of mutational robustness, are shown in Fig. 12(a) and (b), respectively. The mutational robustness at an intermediate temperature observed in MC simulations, as described in the previous section, is thus interpreted as a consequence of the evolution of the fitness landscape at such temperatures.

To confirm this schematic picture of the fitness landscape, we have numerically obtained the fitness distribution of mutated $\boldsymbol{J}$s around the adapted $\boldsymbol{J}$. Here we have computed the fitness values Ψ for $\boldsymbol{J}$s mutated from

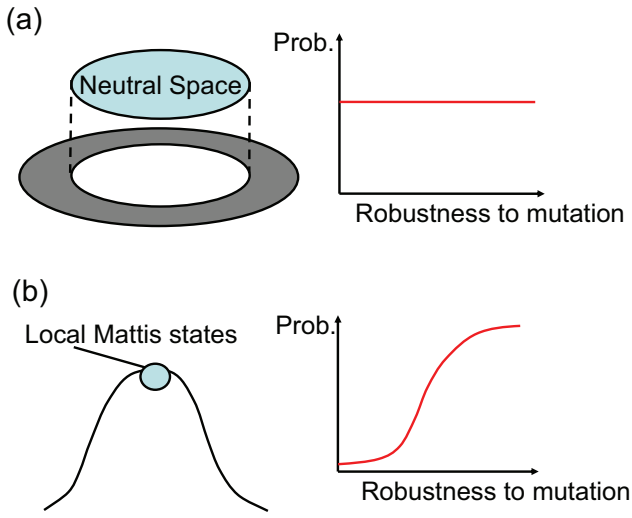

FIG. 12: (color online) Schematic representation of the fitness landscape and the probability distribution of the robustness to mutation estimated from the fitness landscape.
(a) At low T_S, the genotypes that satisfy the condition $W_\Psi(E_0(\boldsymbol{J})) = W(E_0(\boldsymbol{J}))$ exist in the neutral space, where the fitness attains its highest value. The local Mattis states also exist in the neutral space, but they constitute a negligible fraction compared to the frustrated genotypes. The neutral space is surrounded by non-adapted genotypes with sharp boundaries, as in Fig. 11(b). The genotypes around the edge of the neutral space that have very low robustness to mutation survive in an evolutionary sense, similar to the robust genotypes located at the center of neutral space, that are much fewer in number.
(b) For T_S with $T_S^{c1} \le T_S \le T_S^{c2}$, the degeneracy observed at low values of T_S is split, and the local Mattis states have the highest fitness. The fitness of the local Mattis state decreases continuously with increasing mutation. There is a correlation between fitness and robustness, i.e., the local Mattis state with the highest fitness is the most robust to mutation, and the robust genotypes are selected preferentially.

the adapted J with the mutation rate $\mu = 0.1$. In Fig. 13, we have plotted the distribution of fitness values at $T_S = 10^{-3}$ and $T_S = 2.0$ for $N = 15$ and $t = 3$. As shown, the fitness distribution of mutated Js at low T_S has two peaks at $\Psi = 1$ and 0. Hence some mutations are neutral, while others result in a sharp drop in the fitness to its minimal values. In contrast, the fitness of the mutated $\boldsymbol{J}$s at the intermediate T_S is broadly distribute around 0.8. This result supports the schematic fitness landscape in Fig. 12.

At the intermediate T_S, it has been shown that the LMS phase vanishes at sufficiently large t as seen in Fig. 9(c). This could be understood by the T_S and t-dependence of the fitness of the Mattis state. The fitness of the Mattis state at $N = 30$ is plotted in Fig. 14, which is obtained from Eq. (12) and Eq. (14). The fitness of the Mattis state as well as the LMS one at the intermediate temperature region decreases as t is increased. As shown in Fig. 9(a), the fitness value decreases rapidly with T_S

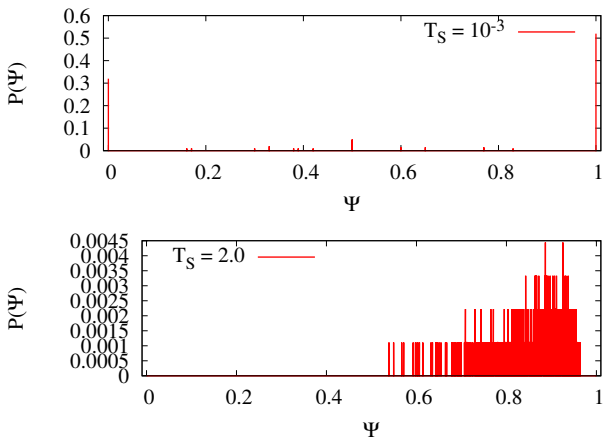

FIG. 13: The fitness distribution of mutated $\boldsymbol{J}$ with mutation rate $\mu = 0.1$, which are generated from the evolved genotypes at $T_S = 10^{-3}$ and at $T_S = 2.0$ at $N = 15$. The distribution is obtained by 100 types of mutated $\boldsymbol{J}$.

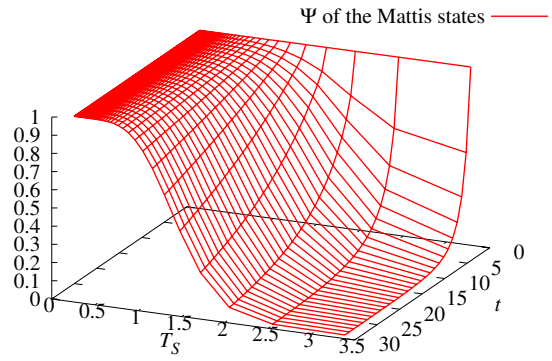

FIG. 14: T_S- and t-dependence of the fitness $\Psi(T_S, t)$ of the Mattis state at $N = 30$.

as t increases. The drop with the increase of T_S is more prominent for larger t, and thus the temperature interval to support the LMS gets narrower with the increase of t, and is expected to disappear for large t.

V. CONCLUSIONS AND DISCUSSIONS

We have considered the evolution of a Hamiltonian system to generate a specific configuration for target spins that captures the basic features required to study the evolution. In this study, we adopted a Markov process, which is given by temperature T_J and fitness $\Psi(\boldsymbol{J})$, for evolutional dynamics. By performing numerical simulation, we found that a specific subset of $\boldsymbol{J}$ with low energy and high fitness is evolved at an intermediate T_S and low T_J. From the statistical-mechanical viewpoint, we focused on frustration and found that the interactions $\boldsymbol{J}$ that evolved at the intermediate T_S are less frustrated. We called these $\boldsymbol{J}$ the local Mattis states. In general, the less frustrated $\boldsymbol{J}$ states are robust to mutation. Hence, the robustness of evolving states to mutation is realized at intermediate temperatures $T_S^{c1} \leq T_S \leq T_S^{c2}$. In other words, robustness to thermal noise introduces mutational robustness; this has also been recently discussed for gene regulation network models[12]. The relevance of thermal noise to robust evolution is thus demonstrated.

The mechanism by which the mutational robustness is achieved could be understood by a statistical-mechanical argument. The T_S-dependence of the fitness landscape was determined by explicitly calculating the fitness $\Psi(\boldsymbol{J})$ for a typical case. It was found that the correlation between fitness and mutational robustness is generated at intermediate T_S, while it disappears at sufficiently low T_S. Hence, the mutationally robust interactions $\boldsymbol{J}$ are obtained as a result of the selection under a certain level of noise. The evolution of the mutational robustness at the intermediate T_S is confirmed by applying statistical-mechanical theory.

We also found that for the interactions evolved at the intermediate T_S, the relaxation to equilibrium progresses smoothly without being stuck at metastable states. For protein folding, such an energy landscape was proposed in terms of the consistency principle by Go [16] and as a funnel-landscape by Onuchic et al. [18]. The funnel energy landscape has no ruggedness around the folded state, in contrast to the spin-glass state. On the other hand, far from the folded state, the ruggedness in the landscape remains; this is different from the global attraction in the ferromagnetic or Mattis state. We note that the landscape in a local Mattis state that is generated as a result of evolution at $T_S^{c1} \leq T_S \leq T_S^{c2}$ is such a funnel landscape, as there is no frustration around the target spins, whereas frustration around non-target spins can result in ruggedness in the landscape far from the target configuration. Indeed, such a smooth and quick relaxation process is observed only for $T_S^{c1} \leq T_S \leq T_S^{c2}$, whereas the relaxation is often stuck at metastable states for a system evolved below T_S^{c1}. On the basis of this correspondence between the funnel landscape and the local Mattis state, we expect that the funnel landscape is characterized by a state with $\Phi_1 = \Phi_2 = 1$.

Biological systems have evolved and functioned at a range of temperatures. Recall that for a Hamiltonian evolved with $T_S < T_S^{c1}$, a large number of time steps is required to reach a spin configuration having the highest fitness, and the number of steps seems to increase with the number of total spins. High fitness is not achieved for a low-temperature region within a biologically acceptable time span, i.e, within a single generation. Accordingly, adaptive evolution is possible only when sufficient thermal noise is present; this corresponds to $T_S > T_S^{c1}$.

We have found that the funnel-like landscape evolves in such biologically relevant temperature regions. In fact,

such a landscape is commonly observed not only in protein folding but also in gene expression dynamics [12, 19] and in the morphogenesis of multicellular organisms [20]. Our result implies that a landscape that allows smooth relaxation dynamics toward the target phenotype, such as the above-mentioned landscape, is realized as a consequence of dynamics that are robust to thermal noise as well as to mutation. We expect that this type of landscape that results from robustness can be found in general in biological evolution. It will not be restricted to Hamiltonian dynamics for protein folding; rather, it will be generally applicable in developmental dynamics. Indeed, recent studies on the evolution of the gene regulation network also demonstrate that the funnel-type dynamics evolve at the intermediate range of noise amplitude values [12, 13]. Our results may explain the ubiquity of such funnel-like dynamics in evolved biological systems.

Although the frustration measure in the present paper is not directly applied to the gene regulation network, the transition to the robust developmental landscape is common. It will be interesting to study the similarities and differences in the transitions in the spin Hamiltonian model presented here and the dissipative gene expression dynamical model.

It is also interesting that there is an optimal number of target spins for achieving the local Mattis state with a funnel energy landscape over a wide range of temperatures. The number of non-target spins $N - t$ controls the redundancy of the system. If t is too small, the target configurations will be more easily perturbed by non-target spins, and fitted states will be less robust to the thermal noise. On the other hand, if it is too large, such $\boldsymbol{J}$ configurations with high fitness are too much limited in the whole configurations, and hence it will not be easily accessed or may be destabilized by mutation. The existence of an appropriate number of redundant spins is important to achieve robustness. Indeed, in proteins, amino-acid residues that are responsible for a function are limited and their number is typically smaller than the number of other proteins. Possible links between redundancy and the evolvability are pointed out in [9], but they still wait to be established quantitatively by theoretical analysis. The present demonstration of the optimal fraction of target elements for realizing robust functions will be important in this context.

In this study, we observed transitions at T_S^{c1} and T_S^{c2}. The phase below T_S^{c1} corresponds to the spin-glass phase and that above T_S^{c2} corresponds to the paramagnetic phase in the context of statistical physics; in contrast, the local Mattis phase between the two phases, which could correspond to the funnel landscape, is a novel discovery in this study. Since a framework of statistical physics of spin systems has been adopted in the present study, the theoretical concepts developed therein, such as replica symmetry breaking, may be applicable in the context developed here to understand this transition. In particular, our model in this study is a variant of spin Hamiltonian systems with two temperatures: one for spin and the other for interactions $\boldsymbol{J}$. Theoretical analysis of such systems[27, 28] will be relevant for the analysis of the local Mattis state and mutational robustness that we have discussed in this paper.

Acknowledgments

This study was partially supported by a Grant-in-Aid for Scientific Research (No.18079004) from MEXT and JSPS Fellows (No.20−10778) from JSPS.

[1] S. Saito, M. Sasai, and T. Yomo, Proc. Natl. Acad. Sci. USA **94**, 11324 (1997).
[2] L. W. Ancel and W. Fontana, J. Exp. Zool. (Mol. Dev. Evol.) **288**, 242 (2000).
[3] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, Science **297**, 1183 (2002).
[4] C. R. Landry, B. Lemos, S. A. Rifkin, W. J. Dickinson, and D. J. Hartl, Science **317**, 118 (2007).
[5] K. Sato, Y. Ito, T. Yomo, and K. Kaneko, Proc. Natl. Acad. Sci. USA **100**, 14086 (2003).
[6] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Rev. Genet. **6**, 451 (2005).
[7] U. Alon , M. G. Surette, N. Barkai, and S. Leibler, Nature (London) **397**, 168 (1999).
[8] S. Ciliberti, O. C. Martin, and A. Wagner, PLoS Comput. Biol. **3**, e15 (2007).
[9] A. Wagner, *Robustness and Evolvability in Living Systems* (Princeton University Press, New Jersey, 2005).
[10] J. Sun and M. W. Deem, Phys. Rev. Lett. **99**, 228107 (2007).
[11] K. Kaneko, *Life: An Introduction to Complex Systems Biology* (Springer-Verlag, Berlin, New York, 2006).
[12] K. Kaneko, PLoS ONE **2**, e434 (2007).
[13] K. Kaneko, Chaos **18**, 026112 (2008).
[14] K. Kaneko, and C. Furusawa, J. Theor. Biol. **240**, 78 (2006).
[15] C. H. Waddington, *The Strategy of the Genes* (George Allen & Unwin LTD, Bristol, 1957).
[16] N. Go, Ann. Rev. Biophys. Bioeng. **12**, 183 (1983).
[17] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 84 (1987).
[18] J. N. Onuchic and P. G. Wolynes, Current Opinion in Structural Biology **14**, 70 (2004).
[19] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, Proc. Natl. Acad. Sci. USA **101**, 4781 (2004).
[20] K. Kaneko, K. Sato, T. Michiue, K. Okabayashi, K. Ohnuma, H. Danno, and M. Asashima, J. Exp. Zool. B **310**, 492 (2008).
[21] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction* (Oxford University Press, New York, 2001).
[22] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Sci. Pub., 1987).
[23] A. Sakata, K. Hukushima, and K. Kaneko, Phys. Rev. Lett. **102**, 148101 (2009).
[24] K. Hukushima and K. Nemoto, J. Phys. Soc. Jpn. **65**, 1604 (1996).
[25] D. C. Mattis, Phys. Lett. **56**, 421 (1976).
[26] E. V. Nimwegen, J. P. Crutchfield, and M. Huynen, Proc. Natl. Acad. Sci. USA **96**, 9716 (1999).
[27] R. W. Penney, A. C. C. Coolen, and D. Sherrington, J. Phys. A: Math. Gen. **26**, 3681 (1993).
[28] V. Dotsenko, S. Franz, and M. Mézard, J. Phys. A: Math. Gen. **27**, 2351 (1994).